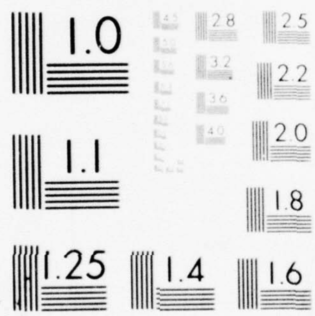
AD-A031 127 CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF STATISTICS F/G 12/1
MAXIMUM LIKELIHOOD ESTIMATION OF STOCHASTIC LINEAR DIFFERENCE E--ETC(U)
JUN 76 G REINSEL N00014-76-C-0930
UNCLASSIFIED TR-112 NL

1 OF 1
AD
A031127

END
DATE
FILMED
11-76



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

Jel 1473

FC.

MAXIMUM LIKELIHOOD ESTIMATION
OF STOCHASTIC LINEAR DIFFERENCE
EQUATIONS WITH AUTOREGRESSIVE
MOVING AVERAGE ERRORS

by

Greg Reinsel

Technical Report No. 112
ONR Report No. 1

June 1976

Research supported by the Office of Naval Research under Contract Number N00014-76-C-0930 (NR 042-367)

Department of Statistics
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

Abstract

A method is proposed for the estimation of a general class of scalar linear time series models. The model takes the form of a stochastic difference equation for the dependent variable with exogenous variable inputs, and the disturbances are autocorrelated through an autoregressive moving average process. In the present paper an asymptotically efficient yet computationally simple estimation procedure (in the time domain) is derived for this model. The resulting estimator is shown to be asymptotically equivalent to the maximum likelihood estimator and to possess a limiting multivariate normal distribution.

1. INTRODUCTION

We consider the estimation of the parameters in the model

$$(1)\qquad y_t - \sum_{i=1}^{r} \alpha_i y_{t-i} = \sum_{i=1}^{k} \beta_i x_{ti} + u_t,$$

$$(2)\qquad u_t - \sum_{i=1}^{p} \phi_i u_{t-i} = \epsilon_t + \sum_{i=1}^{q} \gamma_i \epsilon_{t-i}, \quad (t = \ldots,-1,0,1,\ldots).$$

With the use of the lag operator $\mathcal{L}$ such that $\mathcal{L}^i y_t = y_{t-i}$, the model can also be written as

$$(3)\qquad A(\mathcal{L}) y_t = \sum_{i=1}^{k} \beta_i x_{ti} + u_t,$$

where $A(\mathcal{L}) = 1 - \alpha_1\mathcal{L} - \ldots - \alpha_r\mathcal{L}^r$, and

$$(4)\qquad \phi(\mathcal{L}) u_t = \Gamma(\mathcal{L})\epsilon_t,$$

where $\phi(\mathcal{L}) = 1 - \phi_1\mathcal{L} - \ldots - \phi_p\mathcal{L}^p$, $\Gamma(\mathcal{L}) = 1 + \gamma_1\mathcal{L} + \ldots + \gamma_q\mathcal{L}^q$.

Assumptions used in the estimation of this model are

(i) the ϵ_t are independent and identically distributed with mean 0 and common variance σ^2,

(ii) all roots of $A(z) = 0$, $\phi(z) = 0$, and $\Gamma(z) = 0$ are greater than 1 in absolute value and there are no roots common to the three equations,

(iii) the exogenous variables x_{ti} are nonstochastic sequences which satisfy

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} x_{t+m,i} x_{t+n,j} = \rho_{ij}(n-m) = \rho_{ji}(m-n)$$

exists for $m,n = ,\ldots,-1,0,1,\ldots,$ and $i,j=1,\ldots,k$, with $\rho_{ii}(0) > 0$.

These assumptions imply the following:

(iv) the infinite series $A(z)^{-1} = \sum_{i=0}^{\infty} \lambda_i z^i$, $\emptyset(z)^{-1} = \sum_{i=0}^{\infty} \psi_i z^i$, and

$$\Gamma(z)^{-1} = \sum_{i=0}^{\infty} \delta_i z^i \text{ all converge for } |z| < 1 + \Delta,\ \Delta > 0,$$

(v) the endogenous variable y_t can be expressed as the "steady state solution" to the difference equation (3) as

$$(5) \quad y_t = \sum_{i=1}^{k} \beta_i A(\mathcal{L})^{-1} x_{ti} + A(\mathcal{L})^{-1} u_t = \sum_{i=1}^{k} \sum_{j=0}^{\infty} \beta_i \lambda_j x_{t-j,i} + \sum_{j=0}^{\infty} \lambda_j u_{t-j},$$

while the disturbance u_t can be expressed as the stationary solution to the autoregressive moving average equation (4) as

$$(6) \quad u_t = \emptyset(\mathcal{L})^{-1} \Gamma(\mathcal{L}) \epsilon_t = \sum_{i=0}^{q} \sum_{j=0}^{\infty} \gamma_i \psi_j \epsilon_{t-i-j}, \quad \gamma_0 = 1,$$

(vi) there exists a spectral distribution matrix $F_x(\lambda) = \{F_{mn}(\lambda)\}$, such that $\rho_{mn}(h) = \int_{-\pi}^{\pi} e^{ih\lambda} dF_{mn}(\lambda)$, $(m,n=1,\ldots,k;\ h=0,1,\ldots)$. This can be expressed more compactly as $P(h) = \int_{-\pi}^{\pi} e^{ih\lambda} dF_x(\lambda)$, where $P(h)$ denotes the matrix whose (m,n)th element is $\rho_{mn}(h)$. (See Hannan [14], Chapter 2, or Anderson [3], Chapter 7, for details concerning this assertion.)

Several techniques have previously been introduced for estimating special cases of the above model (3)-(4). For example, when $\beta_1 = \dots = \beta_k = 0$ and $\phi(z) = 1$ the model reduces to

$$A(\mathscr{L})y_t = \Gamma(\mathscr{L})\epsilon_t,$$

the classical time series ARMA model. Early methods of estimation of this model include those proposed by Durbin [9], [10], and Walker [30], [31]. Durbin's method relies on approximating the moving average process $v_t = \epsilon_t + \sum_{i=1}^{q} \gamma_i \epsilon_{t-i}$ by a high-order autoregression, while Walker's procedure maximizes, with respect to the α_i and γ_i, the approximate likelihood function of the first n sample serial correlations of y_t. More recently, Hannan [13], Clevenson [7], and Parzen [24] have constructed estimates based on Fourier transformation of the data and spectral methods. Akaike [1] has shown that Hannan's procedure is approximately a Newton-Raphson method in the frequency domain, while the methods of Clevenson and Parzen are approximations to the method of scoring in the frequency domain (using an alternative parametrization of the moving average process.). The method of Box and Jenkins [5] is to maximize the likelihood function by computing its value at a grid of trial values of the parameters. They also consider nonlinear least squares estimation of the parameters by numerical methods. Anderson [4] estimates the parameters using the method of scoring and Newton-Raphson methods (in the time domain) under the assumption $y_0 = y_{-1} = \dots = y_{1-r} = 0$ and $\epsilon_0 = \epsilon_{-1} = \dots = \epsilon_{1-q} = 0$. The method that will be presented later in this paper for estimating the more general linear model (3)-(4) is related to Anderson's method for this special case.

For the model (3)-(4) containing exogenous variables, methods of estimation have been presented only in special cases. For the difference equation model (3) with pure autoregressive errors in (4), Wallis [33] has suggested an Aitken generalized least squares estimator of the parameters in (3) using an estimate of the covariance matrix of the error term u_t. However, as noted by Amemiya and Fuller [2], Maddala [21] and others, this method is not asymptotically efficient due to the presence of lagged values of the endogenous variable y_t as regressors. Recently Hatanaka [17] has presented an efficient two-step estimation procedure for this model which is identical to the procedure that will be proposed in the present paper for this special case. It will be shown that Hatanaka's procedure is approximately a Newton-Raphson method in the time domain. For the special case of a moving average errors model in (4) with equality of moving average and difference equation coefficients (i.e., the distributed lag model with $\gamma_i = -\alpha_i$), Dhrymes [8] has presented a method of estimation based on Newton-Raphson techniques which is similar to the method to be proposed. Estimation of the model containing a general moving average errors process in (4) has been considered by Hannan and Nicholls [16] using Fourier transformed data. Phillips [25], Trivedi [29], and Hendry and Trivedi [18] have estimated this same model by iterative solution of the maximum likelihood equations, their methods being somewhat similar in this special case to the method to be presented in this paper. An excellent summary of methods of estimation of the difference equation model (3) with moving average disturbances in (4) is provided by Nicholls, Pagan, and Terrell [23]. Box and Jenkins [5] and Pierce [27] have considered the estimation by iterative nonlinear least squares methods of a general transfer function model which is similar but not identical to the model to be considered here.

The purpose of the present paper is to obtain an estimation procedure for the linear time series model (3)-(4) which is asymptotically efficient yet computationally simple. The method to be proposed uses the maximum likelihood approach and is based on Newton-Raphson techniques applied to the likelihood equations. The resulting "Newton-Raphson" estimator is shown to be asymptotically equivalent to the maximum likelihood estimator and to possess a limiting multivariate normal distribution.

2. THE METHOD OF ESTIMATION

For the estimation of the parameters in the model (3)-(4), let us suppose that the observations

$$y_t,\ y_{t-1},\ldots,y_{t-r},\ x_{t1},\ldots,x_{tk}, \quad \text{are available for } t=1,\ldots,T.$$

To motivate the estimation procedure we will assume the ϵ_t are normally distributed and use the maximum likelihood approach. To simplify the form of the likelihood function certain assumptions will be made concerning the initial observations and disturbances. This is necessary due to the dynamic time series structure of the model, and in particular to the complicated form of the inverse of the covariance matrix of consecutive observations from a moving average process. First, we consider the initial observations $y_{1-r},\ldots,y_0,\ldots,y_p$ as fixed, and estimate from the likelihood function conditional on these values. Second, we assume that the initial disturbances $\epsilon_{p+1-q},\ldots,\epsilon_p$ are equal to their unconditional expectations, which are 0. Then introducing the $(T-p)\times(T-p)$ lag matrix L which has 1's on the diagonal directly below the main diagonal and 0's elsewhere, we define the $(T-p)\times(T-p)$ matrix

$$G = I + \sum_{i=1}^{q} \gamma_i L^i .$$

(Note that by condition (iv) of Section 1, $G^{-1} = \sum_{i=0}^{T-p-1} \delta_i L^i$.)

Thus defining the vectors

$$Y = (y_{p+1},\ldots,y_T)', \qquad X_i = (x_{p+1,i},\ldots,x_{Ti})', \ (i=1,\ldots,k),$$

$$U = (u_{p+1},\ldots,u_T)', \qquad \epsilon = (\epsilon_{p+1},\ldots,\epsilon_T)',$$

we can express the entire (modified) model in vector form as

$$(7) \qquad Y - \sum_{i=1}^{r} \alpha_i \mathcal{L}^i Y = \sum_{i=1}^{k} \beta_i X_i + U,$$

$$(8) \qquad U - \sum_{i=1}^{p} \phi_i \mathcal{L}^i U = \epsilon + \sum_{i=1}^{q} \gamma_i L^i \epsilon,$$

where it should be noted that $\mathcal{L}Y = (y_p,\ldots,y_{T-1})'$, $\mathcal{L}U = (u_p,\ldots,u_{T-1})'$, while $L\epsilon = (0, \epsilon_{p+1},\ldots,\epsilon_{T-1})'$. The equations (7)-(8) can be written more compactly as

$$(9) \qquad A(\mathcal{L})Y = \sum_{i=1}^{k} \beta_i X_i + U,$$

(10) $\qquad \phi(\mathcal{L})U = G\epsilon.$

On the assumption of normality of the ϵ_t, the (modified) likelihood function of the observations $y_{p+1},\ldots,y_T$, given $y_{1-r},\ldots,y_p$, is

$$F = \frac{1}{(2\pi)^{\frac{1}{2}(T-p)}(\sigma^2)^{\frac{1}{2}(T-p)}} \exp\left[-\frac{1}{2\sigma^2}(\phi(\mathcal{L})U)'G'^{-1}G^{-1}(\phi(\mathcal{L})U)\right],$$

where U is expressible in terms of the observable quantities Y and X_i through equation (9). Then using the fact that

$$\frac{\partial}{\partial\gamma_m} G^{-1} = -G^{-1}(\frac{\partial}{\partial\gamma_m} G)G^{-1} \quad \text{and} \quad \frac{\partial}{\partial\gamma_m} G = L^m,$$

we obtain the partial derivatives

$$\frac{\partial \log F}{\partial\alpha_m} = \frac{1}{\sigma^2}(\mathcal{L}^m\phi(\mathcal{L})Y)'G'^{-1}G^{-1}(\phi(\mathcal{L})U), \qquad (m=1,\ldots,r),$$

$$\frac{\partial \log F}{\partial\beta_m} = \frac{1}{\sigma^2}(\phi(\mathcal{L})X_m)'G'^{-1}G^{-1}(\phi(\mathcal{L})U), \qquad (m=1,\ldots,k),$$

$$\frac{\partial \log F}{\partial\phi_m} = \frac{1}{\sigma^2}(\mathcal{L}^mU)'G'^{-1}G^{-1}(\phi(\mathcal{L})U), \qquad (m=1,\ldots,p),$$

$$\frac{\partial \log F}{\partial \gamma_m} = \frac{1}{\sigma^2} (\phi(\mathcal{L})U)'G'^{-1}L^{m'}G'^{-1}G^{-1}(\phi(\mathcal{L})U), \qquad (m=1,\ldots,q).$$

Defining the vector

$$\theta = (\alpha_1,\ldots,\alpha_r,\ \beta_1,\ldots,\beta_k,\ \phi_1,\ldots,\phi_p,\ \gamma_1,\ldots,\gamma_q)'$$

and the matrix

$$W = (\mathcal{L}\phi(\mathcal{L})Y,\ldots,\mathcal{L}^r\phi(\mathcal{L})Y,\ \phi(\mathcal{L})X_1,\ldots,\phi(\mathcal{L})X_k,\mathcal{L}U,\ldots,\mathcal{L}^pU,L\epsilon,\ldots,L^q\epsilon),$$

we can express these derivatives in vector form as

$$(11) \qquad \frac{\partial \log F}{\partial \theta} = \frac{1}{\sigma^2}\ W'G'^{-1}G^{-1}(\phi(\mathcal{L})U).$$

Setting these derivatives equal to 0 leads to maximum likelihood equations which are nonlinear in the parameters θ. Thus these equations can only be solved by numerical procedures such as the Newton-Raphson method. The Newton-Raphson method for solving equation (11) is based on the Taylor's expansion

$$\frac{\partial \log F}{\partial \theta} = \left.\frac{\partial \log F}{\partial \theta}\right|_{\theta_0} + \left.\frac{\partial^2 \log F}{\partial\theta\partial\theta'}\right|_{\theta^*} (\theta - \theta_0),$$

where $\|\theta^* - \theta_0\| \leq \|\theta - \theta_0\|$ and $\|\cdot\|$ denotes the usual Euclidean norm. The Newton-Raphson equations for an approximate maximum likelihood estimator $\hat{\theta}$ are

$$(12) \qquad -\left.\frac{\partial^2 \log F}{\partial\theta\partial\theta'}\right|_{\theta_0} (\hat{\theta} - \theta_0) = \left.\frac{\partial \log F}{\partial \theta}\right|_{\theta_0} ,$$

where θ_0 is an initial estimate of θ. Thus the Hessian of log F, $\frac{\partial^2 \log F}{\partial\theta\partial\theta'}$, plays a dominant role in the Newton-Raphson method. It can be shown that an approximation to $-\frac{\partial^2 \log F}{\partial\theta\partial\theta'}$ is given by $\frac{1}{\sigma^2} W'G'^{-1}G^{-1}W$. This approximation involves the omitting of certain terms in the Hessian of log F which, when divided by T, converge to 0 in probability as $T \to \infty$. We will express this "asymptotic" approximation as

$$(13) \qquad -\frac{\partial^2 \log F}{\partial\theta\partial\theta'} \doteq \frac{1}{\sigma^2} W'G'^{-1}G^{-1}W.$$

To obtain the Newton-Raphson estimator of θ, we assume that we have an initial estimate,

$\theta_0 = (\tilde{\alpha}_1,\ldots,\tilde{\alpha}_r, \tilde{\beta}_1,\ldots,\tilde{\beta}_k, \tilde{\phi}_1,\ldots,\tilde{\phi}_p, \tilde{\gamma}_1,\ldots,\tilde{\gamma}_q)'$, which is a consistent estimate of θ to the order $T^{-\frac{1}{2}}$ in probability, i.e., $\theta_0 - \theta = O_p(T^{-\frac{1}{2}})$. This estimate may be obtained as follows:

(a) Obtain consistent estimates $\tilde{\alpha}_1,\ldots,\tilde{\alpha}_r, \tilde{\beta}_1,\ldots,\tilde{\beta}_k$ from equation (7) using the method of instrumental variables estimation. (See Liviatan [20] and Dhrymes [8], Chapter 5). Then compute the residuals

$$\tilde{u}_t = y_t - \sum_{i=1}^{r} \tilde{\alpha}_i y_{t-i} - \sum_{i=1}^{k} \tilde{\beta}_i x_{ti}, \quad (t=1,\ldots,T).$$

(b) Using the calculated residuals $\tilde{u}_t$, we obtain consistent estimates of the autocovariances $\sigma(s) = E(u_t u_{t-s})$, s=0,1,..., of the u_t as

$$c(s) = \frac{1}{T} \sum_{t=s+1}^{T} (\tilde{u}_t - \bar{u})(\tilde{u}_{t-s} - \bar{u}) = c(-s), \quad (s=0,1,\ldots,p+q),$$

where $\bar{u} = \frac{1}{T} \sum_{t=1}^{T} \tilde{u}_t$. We then estimate the parameters ϕ_i consistently by solving the Yule-Walker type equations

$$c(s) - \sum_{i=1}^{p} \tilde{\phi}_i c(s-i) = 0, \qquad (s=q+1,\ldots,q+p).$$

(c) Having obtained the estimates $\tilde{\phi}_i$, we form

$$c_v(s) = \sum_{j=0}^{p} \sum_{\ell=0}^{p} \tilde{\phi}_j \tilde{\phi}_\ell \, c(s+j-\ell) = c_v(-s), \quad s=0,1,\ldots,q, \quad \text{where}$$

$\tilde{\phi}_0 = -1$, and

$$\tilde{f}_v(\lambda) = \frac{1}{2\pi} \sum_{s=-q}^{q} c_v(s) e^{-is\lambda} = \frac{1}{2\pi} [c_v(0) + 2 \sum_{s=1}^{q} c_v(s) \cos s\lambda], \quad -\pi \leq \lambda \leq \pi.$$

These are consistent estimates of the autocovariances and spectral density, respectively, of the moving average process $v_t = \epsilon_t + \sum_{i=1}^{q} \gamma_i \epsilon_{t-i}$. If $\tilde{f}_v(\lambda) \geq 0$ for $-\pi \leq \lambda \leq \pi$, then we may factorize this in the form

$$\tilde{f}_v(\lambda) = \frac{\tilde{\sigma}^2}{2\pi} \left| \sum_{h=0}^{q} \tilde{\gamma}_h e^{ih\lambda} \right|^2 ,$$

where the $\tilde{\gamma}_h$ are real with $\tilde{\gamma}_0 = 1$. The consistent estimates $\tilde{\gamma}_h$ are obtained as the solution to the equations

$$c_v(s) = \tilde{\sigma}^2 \sum_{h=0}^{q-s} \tilde{\gamma}_h \tilde{\gamma}_{h+s} , \quad (h=0,1,\ldots,q),$$

which may be solved using an algorithm of Wilson [35]. Consistent estimates of the γ_h may also be obtained by a method which does not require the factorizing of the spectral density $\tilde{f}_v(\lambda)$ (see Hannan [13]).

Now we let

$$\tilde{A}(\mathcal{L}) = 1 - \tilde{\alpha}_1\mathcal{L} - \dots - \tilde{\alpha}_r\mathcal{L}^r, \quad \tilde{\phi}(\mathcal{L}) = 1 - \tilde{\phi}_1\mathcal{L} - \dots - \tilde{\phi}_p\mathcal{L}^p,$$

$$\tilde{G} = I + \sum_{i=1}^{q} \tilde{\gamma}_i L^i,$$

$$\tilde{U} = \tilde{A}(\mathcal{L})Y - \sum_{i=1}^{k} \tilde{\beta}_i X_i, \qquad \tilde{\epsilon} = \tilde{G}^{-1}\tilde{\phi}(\mathcal{L})\tilde{U} ,$$

$$\tilde{W} = (\mathcal{L}\tilde{\phi}(\mathcal{L})Y, \dots, \mathcal{L}^r\tilde{\phi}(\mathcal{L})Y, \tilde{\phi}(\mathcal{L})X_1, \dots, \tilde{\phi}(\mathcal{L})X_k, \mathcal{L}\tilde{U}, \dots, \mathcal{L}^p\tilde{U}, L\tilde{\epsilon}, \dots, L^q\tilde{\epsilon}).$$

Then using (11) and (13) in equation (12), we obtain the following Newton-Raphson equations for $\hat{\theta}$,

$$\tilde{W}'\tilde{G}'^{-1}\tilde{G}^{-1}\tilde{W}(\hat{\theta} - \theta_0) = \tilde{W}'\tilde{G}'^{-1}\tilde{G}^{-1}(\tilde{\phi}(\mathcal{L})\tilde{U}) .$$

Since

$$\tilde{W}\theta_0 + \tilde{\phi}(\mathcal{L})\tilde{U} = \sum_{i=1}^{r} \tilde{\alpha}_i\mathcal{L}^i\tilde{\phi}(\mathcal{L})Y + \sum_{i=1}^{k} \tilde{\beta}_i\tilde{\phi}(\mathcal{L})X_i + \sum_{i=1}^{p} \tilde{\phi}_i\mathcal{L}^i\tilde{U}$$

$$+ \sum_{i=1}^{q} \tilde{\gamma}_i L^i\tilde{\epsilon} + \tilde{\phi}(\mathcal{L})\tilde{U}$$

$$= \tilde{\phi}(\mathcal{L})(Y - \tilde{U}) + \tilde{U} - \tilde{\epsilon} + \tilde{\phi}(\mathcal{L})\tilde{U}$$

$$= \tilde{\phi}(\mathcal{L})Y + \tilde{U} - \tilde{\epsilon},$$

the Newton-Raphson equations can be written in the form

$$(14) \quad \tilde{W}'\tilde{G}'^{-1}\tilde{G}^{-1}\tilde{W}\theta = \tilde{W}'\tilde{G}'^{-1}\tilde{G}^{-1}(\tilde{\phi}(\mathcal{L})Y + \tilde{U} - \tilde{\epsilon}).$$

The Newton-Raphson solution $\hat{\theta}$ to equation (14) can be interpreted as the generalized least squares solution to the identity

$$(15)\quad \tilde{\phi}(\mathcal{L})Y + \tilde{U} - \tilde{\epsilon}$$

$$= \sum_{i=1}^{r} \alpha_i \mathcal{L}^i \tilde{\phi}(\mathcal{L})Y + \sum_{i=1}^{k} \beta_i \tilde{\phi}(\mathcal{L})X_i + \sum_{i=1}^{p} \phi_i \mathcal{L}^i \tilde{U} + \sum_{i=1}^{q} \gamma_i L^i \tilde{\epsilon} + \tilde{G}\epsilon$$

$$+ \sum_{i=1}^{p} (\tilde{\phi}_i - \phi_i)(\mathcal{L}^i \tilde{U} - \mathcal{L}^i U) + \sum_{i=1}^{q} (\tilde{\gamma}_i - \gamma_i)(L^i \tilde{\epsilon} - L^i \epsilon),$$

where the last two terms on the right hand side of the equation are to be neglected and the error term $\tilde{G}\epsilon = \epsilon + \sum_{i=1}^{q} \tilde{\gamma}_i L^i \epsilon$ is treated as having covariance matrix $\sigma^2 \tilde{G}\tilde{G}'$.

We conclude this section by discussing briefly the computations that are needed to complete the estimation procedure and obtain the estimator $\hat{\theta}$ as the solution to equation (14). Once the initial estimate θ_0 and the residuals $\tilde{u}_t$ have been obtained, we compute

$$\tilde{\phi}(\mathcal{L})y_t = y_t - \sum_{i=1}^{p} \tilde{\phi}_i y_{t-i} \qquad \text{for } t=p+1-r,\ldots,T,$$

and similarly compute $\tilde{\phi}(\mathcal{L})x_{ti}$, $i=1,\ldots,k$, and $\tilde{\phi}(\mathcal{L})\tilde{u}_t$ for $t=p+1,\ldots,T$. Next we obtain the vector $\tilde{\epsilon} = \tilde{G}^{-1}\tilde{\phi}(\mathcal{L})\tilde{U}$ recursively from the equation $\tilde{G}\tilde{\epsilon} = \tilde{\phi}(\mathcal{L})\tilde{U}$ as

$$\tilde{\epsilon}_t = \tilde{\phi}(\mathcal{L})\tilde{u}_t - \sum_{i=1}^{q} \tilde{\gamma}_i \tilde{\epsilon}_{t-i} \qquad \text{for } t=p+1,\ldots,T,$$

where $\tilde{\epsilon}_{p+1-q} = \ldots = \tilde{\epsilon}_p = 0$. Then forming the matrix $\tilde{W}$ as defined above, we compute the columns of the matrix of "independent" variables $\bar{W} = \tilde{G}^{-1}\tilde{W}$ recursively from $\tilde{G}\bar{W} = \tilde{W}$ similar to the computation of $\tilde{\epsilon}$, and we also compute the vector of the "dependent" variable $\bar{Y} = \tilde{G}^{-1}(\tilde{\varnothing}(\mathcal{L})Y + \tilde{U} - \tilde{\epsilon})$ recursively from $\widetilde{G}\bar{Y} = \tilde{\varnothing}(\mathcal{L})Y + \tilde{U} - \tilde{\epsilon}$. Finally, the estimator $\hat{\theta}$ is simply the least squares solution to the regression of $\bar{Y}$ on $\bar{W}$, i.e.,

$$\hat{\theta} = (\bar{W}'\bar{W})^{-1}\bar{W}'\bar{Y} .$$

3. THE ASYMPTOTIC DISTRIBUTION OF THE ESTIMATOR

Since the exact finite sample distribution of the proposed estimator $\hat{\theta}$ is too complicated to be obtained in closed form, we will consider only asymptotic properties of the estimator as $T \to \infty$. To describe the asymptotic distribution of the estimator $\hat{\theta}$ we introduce the matrices M, N, Π, H, Z, K, Ξ, T, and Ω whose (m,n)th elements are defined respectively by

$$\mu_{m-n} = \lim_{T\to\infty} \frac{1}{T}\Big(\sum_{j=1}^{k} \beta_j A(\mathcal{L})^{-1}\mathcal{L}^m \varnothing(\mathcal{L}) X_j\Big)'$$

$$\cdot G'^{-1}G^{-1}\Big(\sum_{j=1}^{k} \beta_j A(\mathcal{L})^{-1}\mathcal{L}^n \varnothing(\mathcal{L}) X_j\Big)$$

$$= \int_{-\pi}^{\pi} \frac{|\varnothing(e^{i\lambda})|^2}{|A(e^{i\lambda})|^2|\Gamma(e^{i\lambda})|^2} e^{i(m-n)\lambda}\, d(\beta' F_x(\lambda)\beta), \quad (m, n=1, \ldots, r),$$

where $\beta = (\beta_1, \ldots, \beta_k)'$,

$$\nu_{mn} = \lim_{T\to\infty} \frac{1}{T}\Big(\sum_{j=1}^{k} \beta_j A(\mathcal{L})^{-1}\mathcal{L}^m \varnothing(\mathcal{L}) X_j\Big)' G'^{-1}G^{-1}(\varnothing(\mathcal{L}) X_n)$$

$$= \sum_{j=1}^{k} \beta_j \int_{-\pi}^{\pi} \frac{|\varnothing(e^{i\lambda})|^2}{A(e^{i\lambda})|\Gamma(e^{i\lambda})|^2} e^{im\lambda}\, dF_{jn}(\lambda),$$

$$(m=1, \ldots, r;\ n=1, \ldots, k),$$

$$\pi_{mn} = \lim_{T\to\infty} \frac{1}{T}(\phi(\mathcal{L})X_m)'G'^{-1}G^{-1}(\phi(\mathcal{L})X_n)$$

$$= \int_{-\pi}^{\pi} \frac{|\phi(e^{i\lambda})|^2}{|\Gamma(e^{i\lambda})|^2}\, dF_{mn}(\lambda), \qquad (m,n=1,\ldots,k),$$

$$\eta_{m-n} = \lim_{T\to\infty} \frac{1}{T} E\left[(A(\mathcal{L})^{-1}\phi(\mathcal{L})\mathcal{L}^m U)'G'^{-1}G^{-1}(A(\mathcal{L})^{-1}\phi(\mathcal{L})\mathcal{L}^n U)\right]$$

$$= \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i(m-n)\lambda}}{|A(e^{i\lambda})|^2}\, d\lambda, \qquad (m,n=1,\ldots,r),$$

$$\zeta_{m-n} = \lim_{T\to\infty} \frac{1}{T} E\left[(A(\mathcal{L})^{-1}\phi(\mathcal{L})\mathcal{L}^n U)'G'^{-1}G^{-1}\mathcal{L}^n U\right]$$

$$= \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i(m-n)\lambda}}{A(e^{i\lambda})\phi(e^{-i\lambda})}\, d\lambda, \qquad (m=1,\ldots,r,\ n=1,\ldots,p),$$

$$\varkappa_{m-n} = \lim_{T\to\infty} \frac{1}{T} E\left[(A(\mathcal{L})^{-1}\phi(\mathcal{L})\mathcal{L}^m U)'G'^{-1}G^{-1}(L^n\epsilon)\right]$$

$$= \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i(m-n)\lambda}}{A(e^{i\lambda})\Gamma(e^{-i\lambda})}\, d\lambda, \qquad (m=1,\ldots,r,\ n=1,\ldots,q),$$

$$\xi_{m-n} = \lim_{T\to\infty} \frac{1}{T} E\left[(\mathcal{L}^m U)'G'^{-1}G^{-1}\mathcal{L}^n U\right]$$

$$= \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i(m-n)\lambda}}{|\phi(e^{i\lambda})|^2}\, d\lambda, \qquad (m,n=1,\ldots,p),$$

$$\tau_{m-n} = \lim_{T\to\infty} \frac{1}{T} E[(L^m\epsilon)' G'^{-1} G^{-1} (L^n\epsilon)]$$

$$= \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i(m-n)\lambda}}{|\Gamma(e^{i\lambda})|^2}\, d\lambda, \qquad (m,n=1,\ldots,q).$$

$$\omega_{m-n} = \lim_{T\to\infty} \frac{1}{T} E[(\mathcal{L}^m U)' G'^{-1} G^{-1} (L^n\epsilon)]$$

$$= \frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i(m-n)\lambda}}{\phi(e^{i\lambda})\Gamma(e^{-i\lambda})}\, d\lambda, \qquad (m=1,\ldots,p;\ n=1,\ldots,q).$$

Then we can state the following

<u>Theorem</u>. For the model (1) and (2) under assumptions (i)-(iii) given in Section 1, let $\hat{\theta}$ denote the estimator of $\theta = (\alpha_1,\ldots,\alpha_r,\ \beta_1,\ldots,\beta_k,\ \phi_1,\ldots,\phi_p,\gamma_1,\ldots,\gamma_q)'$ as obtained from equation (14). Then the distribution of $\sqrt{T}\ (\hat{\theta}-\theta)$ converges to a multivariate normal distribution as $T\to\infty$ with mean vector 0 and covariance matrix equal to

$$(16) \qquad \sigma^2 V^{-1} = \sigma^2 \begin{pmatrix} M+H & N & Z & K \\ N' & \Pi & 0 & 0 \\ Z' & 0 & \Xi & \Omega \\ K' & 0 & \Omega' & T \end{pmatrix}^{-1}.$$

Before discussing the proof of the theorem, we make the following comments:

(a) The matrix V defined by (16) can be seen to equal

$$\lim_{T \to \infty} \frac{1}{T} E(W'G'^{-1}G^{-1}W) = \lim_{T \to \infty} - \frac{\sigma^2}{T} E\left(\frac{\partial^2 \log F}{\partial\theta\partial\theta'}\right) .$$

Thus the asymptotic distribution of $\hat{\theta}$ is identical to the asymptotic distribution of the maximum likelihood estimator based on the assumption of normality of the ϵ_t, so that $\hat{\theta}$ is asymptotically efficient relative to the maximum likelihood estimator when the disturbances are normally distributed.

(b) The proof to be given will show that $\hat{\theta}$ converges to θ in probability as $T \to \infty$.

(c) An asymptotically efficient estimate of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{T-p}(\hat{A}(\mathcal{L})\hat{\phi}(\mathcal{L})Y - \sum_{i=1}^{k} \hat{\beta}_i\hat{\phi}(\mathcal{L})X_i)'\hat{G}'^{-1}\hat{G}^{-1}(\hat{A}(\mathcal{L})\hat{\phi}(\mathcal{L})Y - \sum_{i=1}^{k} \hat{\beta}_i\hat{\phi}(\mathcal{L})X_i),$$

where the "^" denotes that these quantities are to be evaluated at $\hat{\theta}$. The estimator $\hat{\sigma}^2$ will be asymptotically uncorrelated with $\hat{\theta}$. Similarly, the covariance matrix of $\hat{\theta}$ can be estimated by

$$\hat{\sigma}^2(\hat{W}'\hat{G}'^{-1}\hat{G}^{-1}\hat{W})^{-1}.$$

(d) The asymptotic properties of the estimator $\hat{\theta}$ do not require an iterative procedure, only an initial consistent estimate. However, in practice one may want to iterate the least squares solution to (14). In particular we suggest a second iteration so that an asymptotically efficient estimate of the covariance matrix of the estimator is obtained simply as a by-product of the least squares estimation.

(e) As we have mentioned in Section 1, the estimation of various special cases of the model (1)-(2) has been considered by others.

We now compare the estimator proposed in this paper with some of the previously proposed estimators for these special cases. For the time series ARMA model

$$A(\mathcal{L})y_t = \Gamma(\mathcal{L})\epsilon_t,$$

equation (15) reduces to

$$Y + \sum_{i=1}^{q} \tilde{\gamma}_i L^i \tilde{\epsilon} = \sum_{i=1}^{r} \alpha_i \mathcal{L}^i Y + \sum_{i=1}^{q} \gamma_i L^i \tilde{\epsilon} + \tilde{G}\epsilon + \sum_{i=1}^{q} (\tilde{\gamma}_i - \gamma_i)(L^i \tilde{\epsilon} - L^i \epsilon).$$

The generalized least squares estimation procedure which results from this equation is similar to Anderson's (see [4]) Newton-Raphson method except for the treatment of initial values of the y_t, i.e., Anderson uses $L^i Y$ in place of $\mathcal{L}^i Y$. In the distributed lag model

$$A(\mathcal{L})y_t = \sum_{i=1}^{k} \beta_i x_{ti} + A(\mathcal{L})\epsilon_t,$$

equation (15) takes the form

$$Y - \sum_{i=1}^{r} \tilde{\alpha}_i L^i \tilde{\epsilon} = \sum_{i=1}^{r} \alpha_i (\mathcal{L}^i Y - L^i \tilde{\epsilon}) + \sum_{i=1}^{k} \beta_i X_i + \tilde{A}\epsilon - \sum_{i=1}^{r} (\tilde{\alpha}_i - \alpha_i)(L^i \tilde{\epsilon} - L^i \epsilon),$$

where $\tilde{A} = I - \sum_{i=1}^{r} \tilde{\alpha}_i L^i$. The estimation of this equation by generalized least squares is related to a method suggested by Dhrymes [8], Chapter 9. For the general moving average errors model

$$A(\mathcal{L})y_t = \sum_{i=1}^{k} \beta_i x_{ti} + \Gamma(\mathcal{L})\epsilon_t,$$

the identity (15) becomes

$$Y + \sum_{i=1}^{q} \tilde{\gamma}_i L^i \tilde{\epsilon} = \sum_{i=1}^{r} \alpha_i \mathcal{L}^i Y + \sum_{i=1}^{k} \beta_i X_i + \sum_{i=1}^{q} \gamma_i L^i \tilde{\epsilon} + \tilde{G}\epsilon$$

$$+ \sum_{i=1}^{q} (\tilde{\gamma}_i - \gamma_i)(L^i \tilde{\epsilon} - L^i \epsilon).$$

The estimation of this identity by generalized least squares is similar to the method of Phillips [25] except for the treatment of the values of the initial disturbances ϵ_t, i.e., Phillips considers these values as parameters to be estimated. Finally, for the pure autoregressive errors model

$$A(\mathcal{L})y_t = \sum_{i=1}^{k} \beta_i x_{ti} + u_t, \qquad \phi(\mathcal{L})u_t = \epsilon_t,$$

Hatanaka [17] has suggested a method identical to the least squares estimation of the identity

$$\tilde{\phi}(\mathcal{L})Y + \sum_{i=1}^{p} \tilde{\phi}_i \mathcal{L}^i \tilde{U} = \sum_{i=1}^{r} \alpha_i \mathcal{L}^i \tilde{\phi}(\mathcal{L})Y + \sum_{i=1}^{k} \beta_i \tilde{\phi}(\mathcal{L})X_i + \sum_{i=1}^{p} \phi_i \mathcal{L}^i \tilde{U} + \epsilon$$

$$+ \sum_{i=1}^{p} (\tilde{\phi}_i - \phi_i)(\mathcal{L}^i \tilde{U} - \mathcal{L}^i U),$$

which is simply (15) in this special case.

<u>Proof of Theorem</u>: We shall not go into great detail here but merely give an outline of the proof. First, we can ignore the effect of the modification of the initial disturbances ϵ_t, and hence the use of the lag matrix G in place of the lag operator $\Gamma(\mathcal{L})$, since the modification has a negligible effect as $T \to \infty$ and the asymptotic properties of the estimator are not affected by this modification

(see Anderson [4]). Then using the (modified) identity (15) and equation (14), we have

$$(17)\quad \hat{\theta} = (\tilde{W}'\tilde{G}'^{-1}\tilde{G}^{-1}\tilde{W})^{-1}\, \tilde{W}'\tilde{G}'^{-1}\tilde{G}^{-1}\, (\tilde{\phi}(\mathcal{L})Y + \tilde{U} - \tilde{\epsilon})$$

$$= (\tilde{W}'\tilde{G}'^{-1}\tilde{G}^{-1}\tilde{W})^{-1}\, \tilde{W}'\tilde{G}'^{-1}\tilde{G}^{-1}\, (\tilde{W}\theta + \tilde{G}\epsilon$$

$$+ \sum_{i=1}^{p} (\tilde{\phi}_i - \phi_i)(\mathcal{L}^i\tilde{U} - \mathcal{L}^i U) + \sum_{i=1}^{q} (\tilde{\gamma}_i - \gamma_i)(L^i\tilde{\epsilon} - L^i\epsilon))$$

It follows that

$$(18)\quad \sqrt{T}\,(\hat{\theta} - \theta) = (\frac{1}{T}\tilde{W}'\tilde{G}'^{-1}\tilde{G}^{-1}\tilde{W})^{-1} \cdot \frac{1}{\sqrt{T}}\tilde{W}'\tilde{G}'^{-1}\epsilon$$

$$+ (\frac{1}{T}\tilde{W}'\tilde{G}'^{-1}\tilde{G}^{-1}\tilde{W})^{-1} \cdot (\sum_{i=1}^{p} \sqrt{T}\,(\tilde{\phi}_i - \phi_i)\frac{1}{T}\tilde{W}'\tilde{G}'^{-1}\tilde{G}^{-1}(\mathcal{L}^i\tilde{U} - \mathcal{L}^i U)$$

$$+ \sum_{i=1}^{q} \sqrt{T}\,(\tilde{\gamma}_i - \gamma_i)\frac{1}{T}\tilde{W}'\tilde{G}'^{-1}\tilde{G}^{-1}(L^i\tilde{\epsilon} - L^i\epsilon)).$$

Now each of the terms $\sqrt{T}\,(\tilde{\phi}_i - \phi_i)\cdot \frac{1}{T}\tilde{W}'\tilde{G}'^{-1}\tilde{G}^{-1}(\mathcal{L}^i\tilde{U} - \mathcal{L}^i U)$ on the right hand side of (18) has a probability limit equal to 0 as $T \to \infty$. This is true since $\sqrt{T}(\tilde{\phi}_i - \phi_i)$ is bounded in probability as $T \to \infty$ by the consistency of $\tilde{\phi}_i$, while $\frac{1}{T}\tilde{W}'\tilde{G}'^{-1}\tilde{G}^{-1}(\mathcal{L}^i\tilde{U} - \mathcal{L}^i U)$ converges to 0 in probability as $T \to \infty$, again by consistency of the initial estimates. The same argument also applies to the terms involving $\tilde{\gamma}_i$ on the right side of (18). Hence we can conclude that the entire second term on the right hand side of (18) converges to 0 in probability as $T \to \infty$, since we will show that the matrix $(\frac{1}{T}\tilde{W}'\tilde{G}'^{-1}\tilde{G}^{-1}\tilde{W})^{-1}$ has a finite probability limit as $T \to \infty$. It also follows by the consistency of the initial estimate θ_0 that if the matrix $\frac{1}{T} W'G'^{-1}G^{-1}W$ possesses a finite probability limit as $T \to \infty$, then the matrix $\frac{1}{T} W'G'^{-1}G^{-1}W\big|_{\theta_0} = \frac{1}{T}\tilde{W}'\tilde{G}'^{-1}\tilde{G}^{-1}\tilde{W}$ will have the same probability limit. And finally, the limiting distribution of the vector $\frac{1}{\sqrt{T}}\tilde{W}'\tilde{G}'^{-1}\epsilon$ will be the same as that of $\frac{1}{\sqrt{T}} W'G'^{-1}\epsilon$, since the difference between the two vectors, $\frac{1}{\sqrt{T}}(\tilde{G}^{-1}\tilde{W} - G^{-1}W)'\epsilon$, converges to 0 in probability as $T \to \infty$. This follows from the fact that all the elements in the vector of differences above are essentially in the form of products of two terms, one of which involves $\sqrt{T}$ times elements of the difference $(\theta_0 - \theta)$ and the other of which involves elements of the vector $\frac{1}{T} W'G'^{-1}\epsilon$. Then since $\underset{T\to\infty}{\text{plim}} \frac{1}{T} W'G'^{-1}\epsilon = 0$ and θ_0 is a consistent estimate of θ to the order $T^{-\frac{1}{2}}$ in probability, arguments similar to those used for the other terms in equation (18) can also be applied to the vector of differences $\frac{1}{\sqrt{T}}(\tilde{G}^{-1}\tilde{W} - G^{-1}W)'\epsilon$. Thus we see from the above arguments that the limiting distribution of $\sqrt{T}(\hat{\theta} - \theta)$ will be identical to that of $(\frac{1}{T} W'G'^{-1}G^{-1}W)^{-1}\frac{1}{\sqrt{T}} W'G'^{-1}\epsilon$. Hence it follows that the results of the Theorem will be established once we show that

(I) $\underset{T\to\infty}{\text{plim}} \frac{1}{T} W'G'^{-1}G^{-1}W = \underset{T\to\infty}{\lim} \frac{1}{T} E(W'G'^{-1}G^{-1}W) = V$, where the matrix V is defined by (16), and

(II) $\frac{1}{\sqrt{T}} W'G'^{-1}\epsilon$ has a limiting normal distribution with mean vector 0 and covariance matrix equal to $\sigma^2 V$.

Proof of (I). We consider the probability limit of a typical element of the matrix $\frac{1}{T} W'G'^{-1}G^{-1}W$. For example, using equations (9), (10), and (5), the fact that $E(x_{ti}\epsilon_s) = 0$ for all $t,s=\ldots,-1,0,1,\ldots$ and $i=1,\ldots,k$, we have

$$(19)\quad \lim_{T\to\infty} \frac{\sigma^2}{T} E\left(-\frac{\partial^2 \log F}{\partial\gamma_n \partial\alpha_m}\right)$$

$$= \lim_{T\to\infty} \frac{1}{T} E[(\mathcal{L}^m \phi(\mathcal{L})Y)'G'^{-1}G^{-1}(L^n\epsilon)]$$

$$= \lim_{T\to\infty} \frac{1}{T} E[(\sum_{i=1}^{k} \beta_i \mathcal{L}^m A(\mathcal{L})^{-1}\phi(\mathcal{L})X_i + \mathcal{L}^m A(\mathcal{L})^{-1}\phi(\mathcal{L})U)'$$

$$\cdot G'^{-1}G^{-1}(L^n\epsilon)]$$

$$= \lim_{T\to\infty} \frac{1}{T} E[(\mathcal{L}^m A(\mathcal{L})^{-1}\varepsilon)'(G^{-1}L^n\epsilon)]$$

$$= \lim_{T\to\infty} \frac{1}{T} E[\sum_{t=p+n+1}^{T} \sum_{u=0}^{t-p-n-1} \sum_{v=0}^{\infty} \delta_u \lambda_v \epsilon_{t-n-u}\epsilon_{t-m-v}$$

$$= \lim_{T\to\infty} \frac{\sigma^2}{T} \sum_{t=p+n+1}^{T} \begin{cases} \sum_{u=0}^{t-p-n-1} \delta_u \lambda_{u+n-m}, & \text{for } m \le n \\ \sum_{u=0}^{t-p-n-1} \delta_{u+m-n}\lambda_u, & \text{for } m \ge n \end{cases}$$

$$= \lim_{T\to\infty} \frac{\sigma^2}{T} \begin{cases} \sum_{u=0}^{T-p-n-1} (T-n-p-u)\, \delta_u \lambda_{u+n-m}, & \text{for } m \le n \\ \sum_{u=0}^{T-p-n-1} (T-n-p-u)\, \delta_{u+m-n} \lambda_u, & \text{for } m \ge n \end{cases}$$

$$= \sigma^2 \begin{cases} \sum_{u=0}^{\infty} \delta_u \lambda_{u+n-m}, & \text{for } m \le n \\ \sum_{u=0}^{\infty} \delta_{u+m-n} \lambda_u, & \text{for } m \ge n \end{cases}$$

$$= \frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i(m-n)\lambda}}{A(e^{i\lambda})\Gamma(e^{-i\lambda})}\, d\lambda$$

$$= \varkappa_{m-n}, \quad (m=1,\ldots,r;\ n=1,\ldots,q).$$

In obtaining the above limit, we have used the fact that the Cesaro summation of a convergent series converges to that sum (see Anderson [3], Lemma 8.3.1). Also, to obtain $\varkappa_{m-n}$ as the probability limit of the quantity in (19), first we have

$$(20) \quad E[\tfrac{1}{T}(\mathcal{L}^m A(\mathcal{L})^{-1} \emptyset(\mathcal{L}) X_i)' G'^{-1} G^{-1} (L^n \boldsymbol{\epsilon})]^2$$

$$= \frac{1}{T^2} E[\sum_{t=p+n+1}^{T} \sum_{u=0}^{t-p-n-1} \sum_{v=0}^{t-p-1} \delta_u \delta_v A(\mathcal{L})^{-1} \emptyset(\mathcal{L}) x_{t-m-v,i} \epsilon_{t-n-u}]^2$$

$$= \frac{1}{T^2} \sum_{t,s=p+n+1}^{T} \sum_{u=0}^{t-p-n-1} \sum_{j=0}^{s-p-n-1} \sum_{v=0}^{t-p-1} \sum_{\ell=0}^{s-p-1} \delta_u \delta_j \delta_v \delta_\ell$$

$$\cdot A(\mathcal{L})^{-1}\phi(\mathcal{L})x_{t-m-v,i}A(\mathcal{L})^{-1}\phi(\mathcal{L})x_{s-m-\ell,i}E(\epsilon_{t-n-u}\epsilon_{s-n-j})$$

$$= \frac{\sigma^2}{T^2}\sum_{t=p+n+1}^{T}\sum_{u=0}^{t-p-n-1}\sum_{j=0}^{T-t+u}\sum_{v=0}^{t-p-1}\sum_{\ell=0}^{t-u+j-p-1}\delta_u\delta_j\delta_v\delta_\ell$$

$$\cdot A(\mathcal{L})^{-1}\phi(\mathcal{L})x_{t-m-v,i}A(\mathcal{L})^{-1}\phi(\mathcal{L})x_{t-u+j-m-\ell,i}\ .$$

Hence (20) is less than

$$\frac{\sigma^2}{T^2}\max_{p+1\leq t\leq T}|A(\mathcal{L})^{-1}\phi(\mathcal{L})x_{t-m,i}|^2\sum_{t=p+n+1}^{T}\left(\sum_{u=0}^{\infty}|\delta_u|\right)^4$$

$$\leq \frac{\sigma^2}{T}\max_{p+1\leq t\leq T}|A(\mathcal{L})^{-1}\phi(\mathcal{L})x_{t-m,i}|^2\left(\sum_{u=0}^{\infty}|\delta_u|\right)^4,$$

which goes to 0 as $T\to\infty$ because of conditions (ii) and (iii) of Section 1 (see also Anderson [3], Lemma 2.6.1). Thus it follows by Tchebychev's inequality that the first of the two terms in (19) converges to 0 in probability as $T\to\infty$. The second term in (19) is

$$\text{(21)}\qquad \frac{1}{T}(\mathcal{L}^m A(\mathcal{L})^{-1}\epsilon)'(G^{-1}L^n\epsilon) = \frac{1}{T}\sum_{t=p+n+1}^{T}\sum_{u=0}^{t-p-n-1}\sum_{v=0}^{\infty}\delta_u\lambda_v\epsilon_{t-n-u}\epsilon_{t-m-v}\ .$$

Now for fixed $v = 0,1,\ldots,$

$$E\left|\frac{1}{T}\sum_{t=p+n+1}^{T}\sum_{u=0}^{t-p-n-1}\delta_u\epsilon_{t-n-u}\epsilon_{t-m-v}\right|\leq\frac{1}{T}\sum_{t=p+n+1}^{T}\sum_{u=0}^{t-p-n-1}$$

$$|\delta_u|E|\epsilon_{t-n-u}\epsilon_{t-m-v}|$$

$$\leq \sigma^2\sum_{u=0}^{\infty}|\delta_u|,$$

so that

$$(22)\quad \lim_{s\to\infty} E\left|\frac{1}{T}\sum_{v=s+1}^{\infty}\sum_{t=p+n+1}^{T}\sum_{u=0}^{t-p-n-1}\lambda_v\delta_u\epsilon_{t-n-u}\epsilon_{t-m-v}\right|$$

$$\leq \lim_{s\to\infty}\sum_{v=s+1}^{\infty}|\lambda_v|\cdot(\sigma^2\sum_{u=0}^{\infty}|\delta_u|) = 0,$$

uniformly in T. Then Markov's inequality, $P(|X|\geq\epsilon)\leq\frac{E|X|}{\epsilon}$ for any random variable X and any $\epsilon > 0$, implies that the term in (22) converges to 0 in probability as $s\to\infty$ uniformly in T. Also, the quantities

$$\frac{1}{T}\sum_{t=p+n+1}^{T}\sum_{u=0}^{t-p-n-1}\delta_u\epsilon_{t-n-u}\epsilon_{t-m-v} = \frac{1}{T}\sum_{u=0}^{T-p-n-1}\sum_{t=p+n+u+1}^{T}\delta_u\epsilon_{t-n-u}\epsilon_{t-m-}$$

have probability limits as $T\to\infty$ equal to $\sigma^2\delta_{v+m-n}$, for $v=\max(0,n-m),\ldots$. Thus it follows from this last result and (22) (see also Anderson [3], Theorem 7.7.1) that the second term in (19),

$$\frac{1}{T}\sum_{t=p+n+1}^{T}\sum_{u=0}^{t-p-n-1}\sum_{v=0}^{\infty}\delta_u\lambda_v\epsilon_{t-n-u}\epsilon_{t-m-v}$$

$$=\frac{1}{T}\sum_{v=0}^{s}\sum_{t=p+n+1}^{T}\sum_{u=0}^{t-p-n-1}\delta_u\lambda_v\epsilon_{t-n-u}\epsilon_{t-m-v}$$

$$+\frac{1}{T}\sum_{v=s+1}^{\infty}\sum_{t=p+n+1}^{T}\sum_{n=0}^{t-p-n-1}\delta_u\lambda_v\epsilon_{t-n-u}\epsilon_{t-m-v}$$

converges in probability as $T \to \infty$ to

$$\sigma^2 \begin{cases} \lim_{s \to \infty} \sum_{v=0}^{s} \delta_v \lambda_{v+n-m}, & \text{for } m \leq n \\ \lim_{s \to \infty} \sum_{v=0}^{s} \delta_{v+m-n} \lambda_v, & \text{for } m \geq n, \end{cases}$$

which is just $\varkappa_{m-n}$ as given in (19). (The argument used here is similar to that given in the proof of Theorem 1 of Hannan and Heyde [15], page 2060). The same type of argument may be used to establish the probability limits for the other elements of the matrix

$$\frac{1}{T} W' G'^{-1} G^{-1} W.$$

Proof of (II): To establish the asymptotic normality of $\frac{1}{\sqrt{T}} W' G'^{-1} \epsilon$, we consider the asymptotic behavior of a single component. For example, again using (9), (10), and (5), we have

$$(23) \quad \frac{\sigma^2}{\sqrt{T}} \frac{\partial \log F}{\partial \alpha_m} = \frac{1}{\sqrt{T}} (\mathcal{L}^m \phi(\mathcal{L}) Y)' G'^{-1} \epsilon$$

$$= \frac{1}{\sqrt{T}} [(\sum_{i=1}^{k} \beta_i \mathcal{L}^m A(\mathcal{L})^{-1} \phi(\mathcal{L}) X_i)' G'^{-1} \epsilon + (\mathcal{L}^m A(\mathcal{L})^{-1} \epsilon)' \epsilon]$$

$$= \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} [(\sum_{u=0}^{t-p-1} \sum_{i=1}^{k} \delta_u \beta_i A(\mathcal{L})^{-1} \phi(\mathcal{L}) x_{t-m-u,i}) \epsilon_t$$

$$+ A(\mathcal{L})^{-1} \epsilon_{t-m} \epsilon_t]$$

$$= \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} [(\sum_{u=0}^{t-p-1} \sum_{i=1}^{k} \delta_u \beta_i A(\mathcal{L})^{-1} \phi(\mathcal{L}) x_{t-m-u,i}) \epsilon_t$$

$$+ \sum_{u=0}^{n} \lambda_u \epsilon_{t-m-u} \epsilon_t]$$

$$+ \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \sum_{u=n+1}^{\infty} \lambda_u \epsilon_{t-m-u} \epsilon_t$$

$$\equiv Z_{Tn} + R_{Tn},$$

where

$$(24) \qquad Z_{Tn} = \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} W_{tn}$$

$$(25) \qquad W_{tn} = (\sum_{u=0}^{t-p-1} \sum_{i=1}^{k} \delta_u \beta_i A(\mathcal{L})^{-1} \phi(\mathcal{L}) x_{t-m-u,i}) \epsilon_t + \sum_{u=0}^{n} \lambda_u \epsilon_{t-m-u} \epsilon_t$$

$$(t=p+1,\ldots,T),$$

and

$$(26) \qquad R_{Tn} = \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} \sum_{u=n+1}^{\infty} \lambda_u \epsilon_{t-m-u} \epsilon_t .$$

Now for all $T > p$,

$$(27) \quad E(R_{Tn}^2) = \frac{1}{T} \sum_{t,s=p+1}^{T} \sum_{u,v=n+1}^{\infty} \lambda_u \lambda_v \, E(\epsilon_{t-m-u} \epsilon_t \epsilon_{s-m-v} \epsilon_s)$$

$$= \frac{\sigma^4}{T} \sum_{t=p+1}^{T} \sum_{u=n+1}^{\infty} \lambda_u^2$$

$$\leq \sigma^4 \sum_{u=n+1}^{\infty} \lambda_u^2 \equiv M_n,$$

and $\lim_{n \to \infty} M_n = 0$ since $\sum_{u=0}^{\infty} \lambda_u^2$ converges. For fixed n, W_{tn} has mean 0 and variance equal to

$$\sigma_{tn}^2 = \left(\sum_{u=0}^{t-p-1} \sum_{i=1}^{k} \delta_u \beta_i A(\mathcal{L})^{-1} \emptyset(\mathcal{L}) x_{t-m-u,i}\right)^2 \sigma^2 + \sigma^4 \sum_{u=0}^{n} \lambda_u^2 .$$

The covariance between W_{tn} and W_{sn}, $t \neq s$, is 0 and

$$E(W_{tn}^2 | \epsilon_{t-1}, \ldots) = \left(\sum_{u=0}^{t-p-1} \sum_{i=1}^{k} \delta_u \beta_i A(\mathcal{L})^{-1} \emptyset(\mathcal{L}) x_{t-m-u,i}\right)^2 \sigma^2$$

$$+ \sigma^2 \sum_{u,v=0}^{n} \lambda_u \lambda_v \epsilon_{t-m-u} \epsilon_{t-m-v}$$

$$+ 2\sigma^2 \left(\sum_{u=0}^{t-p-1} \sum_{i=1}^{k} \delta_u \beta_i A(\mathcal{L})^{-1} \emptyset(\mathcal{L}) x_{t-m-u,i}\right) \sum_{u=0}^{n} \lambda_u \epsilon_{t-m-u} .$$

Thus letting $V_{Tn}^2 = \frac{1}{T} \sum_{t=p+1}^{T} E(W_{tn}^2 | \epsilon_{t-1}, \ldots)$ and using arguments similar to those given in the proof of (I), we can see that as $T \to \infty$ V_{Tn}^2 converges in probability to

$$(28) \quad \sigma_n^2 = \lim_{T \to \infty} E(Z_{Tn}^2)$$

$$= \lim_{T \to \infty} E(V_{Tn}^2)$$

$$= \lim_{T \to \infty} \frac{\sigma^2}{T} \sum_{t=p+1}^{T} \left(\sum_{u=0}^{t-p-1} \sum_{i=1}^{k} \delta_u \beta_i A(\mathcal{L})^{-1} \emptyset(\mathcal{L}) x_{t-m-u,i}\right)^2 + \sigma^4 \sum_{u=0}^{n} \lambda_u^2$$

$$= \sigma^2 \mu_0 + \sigma^4 \sum_{u=0}^{n} \lambda_u^2 ,$$

where u_0 appears in the matrix V defined in (16).

Also, $Z_{Tn} = \frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} W_{tn}$ is a martingale which satisfies a Lindeberg condition

$$(29) \quad \frac{1}{T\, E(V_{Tn}^2)} \sum_{t=p+1}^{T} E[W_{tn}^2 \cdot I(|W_{tn}| \geq \epsilon\sqrt{T\, E(V_{Tn}^2)}\ )] \to 0 \text{ as } T \to \infty$$

for any $\epsilon > 0$, where $I(\cdot)$ denotes the indicator function. Condition (29) can be shown to hold by the use of the same argument as given in the proof of Theorem 2, page 2063, in Hannan and Heyde [15]. (See also the proof of Theorem 2.6.1 in Anderson [3]). Thus through equations (28) and (29), Z_{Tn} satisfies conditions (1) and (2) of Brown [6]. It follows by Theorem 2 given there that Z_{Tn} has a limiting normal distribution as $T \to \infty$, with mean 0 and variance σ_n^2. Finally, using Theorem 7.7.1 in Anderson [3] and the result following equation (27), we can conclude that

$$\frac{\sigma^2}{\sqrt{T}} \frac{\partial \log F}{\partial \alpha_m} = \frac{1}{\sqrt{T}} (\mathcal{L}^m \emptyset(\mathcal{L})Y)'G'^{-1}\epsilon = Z_{Tn} + R_{Tn}$$

has a limiting normal distribution as $T \to \infty$ with mean 0 and variance equal to

$$\sigma^2\mu_0 + \lim_{n \to \infty} \sigma^4 \sum_{u=0}^{n} \lambda_u^2 = \sigma^2\mu_0 + \sigma^4 \sum_{u=0}^{\infty} \lambda_u^2 = \sigma^2 u_0 + \sigma^2 \eta_0 .$$

The asymptotic normality of all other elements of $\frac{1}{\sqrt{T}} W'G'^{-1}\epsilon$ can be obtained in the same manner. A similar argument can also be used to show that the limiting distribution of an arbitrary linear combination of the elements of $\frac{1}{\sqrt{T}} W'G'^{-1}\epsilon$,

$\frac{1}{\sqrt{T}} C'W'G'^{-1}\epsilon$, with $C = (c_1, \ldots, c_{r+k+p+q})'$ an

arbitrary constant vector, is normal with mean 0 and variance $\sigma^2 C'VC$, where V is the matrix defined by (16). Then using the continuity theorem for characteristic functions , we see that

$\frac{1}{\sqrt{T}} W'G'^{-1}\epsilon = \frac{\sigma^2}{\sqrt{T}} \frac{\partial \log F}{\partial \theta}$ has a limiting multivariate normal distribution

$N(0,\sigma^2 V)$ as $T \to \infty$, and thus the theorem is established.

REFERENCES

[1] Akaike, H. "Maximum likelihood identification of Gaussian autoregressive moving average models." Biometrika 60, (1973), 255-265.

[2] Amemiya, T. and Fuller, W. "A comparative study of alternative estimators in a distributed lag model. Econometrica 35, (1967), 509-529.

[3] Anderson, T. W. The Statistical Analysis of Time Series. New York: Wiley, 1971.

[4] Anderson, T. W. "Maximum likelihood estimation of parameters of autoregressive processes with moving average residuals and other covariance matrices with linear structure. The Annals of Statistics 3, (1975), 1283-1304.

[5] Box, G.E.P. and Jenkins, G. M. Time Series Analysis Forecasting and Control. San Francisco: Holden-Day Inc., 1970.

[6] Brown, B.M. "Martingale central limit theorems." The Annals of Mathematical Statistics 42, (1971), 59-66.

[7] Clevenson, M.L. "Asymptotically efficient estimates of the parameters of a moving average time series. Technical Report No. 15, Department of Statistics, Stanford University, (1970).

[8] Dhrymes, P.J. Distributed Lags: Problems of Estimation and Formulation. San Francisco: Holden-Day Inc., 1971.

[9] Durbin, J. "Efficient estimation of parameters in moving average model. Biometrika 46, (1959), 306-316.

[10] Durbin, J. "The fitting of time-series models. Review of the International Statistical Institute 28, (1960), 233-244.

[11] Grether, D. M. and Maddala, G. S. "On the asymptotic properties of some two-step procedures for estimating distributed lag models. International Economic Review 13, (1972), 737-744.

[12] Gupta, Y.P. "Least squares variant of the Dhrymes two-step estimation procedure of the distributed lag model." International Economic Review 10, (1969), 112-113.

[13] Hannan, E. J. "The estimation of mixed moving average autoregressive systems. Biometrika 56, (1969), 579-593.

[14] Hannan, E. J. Multiple Time Series. New York: John Wiley and Sons, Inc., 1970.

[15] Hannan, E. J. and Heyde, C. C. "On limit theorems for quadratic functions of discrete time series." The Annals of Mathematical Statistics 43, (1972), 2058-2066.

[16] Hannan, E. J. and Nicholls, D. F. "The estimation of mixed regression, autoregression, moving average and distributed lag models." Econometrica 40, (1972), 529-548.

[17] Hatanaka, M. "An efficient two-step estimator for the dynamic adjustment model with autoregressive errors." Journal of Econometrics 2, (1974), 199-220.

[18] Hendry, D. F. and Trivedi, P. K. "Maximum likelihood estimation of difference equations with moving average errors: a simulation study. Review of Economic Studies 39, (1972), 117-145.

[19] Kashyap, R. L. "Maximum likelihood identification of stochastic linear systems. I.E.E.E. Transactions on Automatic Control AC-15, (1970), 25-34.

[20] Liviatan, N. "Consistent estimation of distributed lags." International Economic Review 4, (1963), 44-52.

[21] Maddala, G. S. "Generalized least squares with an estimated variance covariance matrix." Econometrica 39, (1971), 23-34.

[22] Mann, H. B. and Wald, A. "On the statistical treatment of linear stochastic difference equations." Econometrica 11, (1943), 173-220.

[23] Nicholls, D. F. Pagan, A. R. and Terrell, R. D. "The estimation and use of models with moving average disturbance terms: a survey." International Economic Review 16, (1975), 113-134.

[24] Parzen, E. "Efficient estimation of stationary time series mixed schemes." Bulletin of the International Statistical Institute 44, (1971), 315-319.

[25] Phillips, A. W. "Estimation of stochastic difference equations with moving average disturbances." Paper delivered to the Econometric Society Meeting, San Francisco, (1966).

[26] Pierce, D. A. "Least squares estimation in the regression model with autoregressive-moving average errors." Biometrika 58, (1971), 299-312.

[27] Pierce, D. A. "Least squares estimation in dynamic-disturbance time series models." Biometrika 59, (1972), 73-78.

[28] Taylor, L. D. and Wilson, T. A. "Three-pass least squares: a method for estimating models with a lagged dependent variable." Review of Economics and Statistics 46, (1964), 329-346.

[29] Trivedi, P. K. "Inventory behavior in U.K. manufacturing 1956-1967." Review of Economic Studies 37, (1970), 517-536.

[30] Walker, A. M. "Large-sample estimation for moving-average models. Biometrika 48, (1961), 343-357.

[31] Walker, A. M. "Large sample estimation of parameters for autoregressive processes with moving-average residuals." Biometrika 49, (1962), 117-131.

[32] Walker, A. M. "Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series." Journal of the Australian Mathematical Society 4, (1964), 363-384.

[33] Wallis, K. F. "Lagged dependent variables and serially correlated errors: a reappraisal of three-pass least squares." Review of Economics and Statistics 49, (1967), 555-567.

[34] Wallis, K. F. "The efficiency of the two-step estimator." Econometrica 40, (1972), 769-770.

[35] Wilson, G. J. "Factorization of the generating function of a pure moving average process." SIAM Journal of Numerical Analysis 6, (1969), 1-7.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE — READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER ONR Report No. 1	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 MAXIMUM LIKELIHOOD ESTIMATION OF STOCHASTIC LINEAR DIFFERENCE EQUATIONS WITH AUTOREGRESSIVE MOVING AVERAGE ERRORS,		5. TYPE OF REPORT & PERIOD COVERED 9 Technical Report,
		6. PERFORMING ORG. REPORT NUMBER Technical Report No. 112
7. AUTHOR(s) 10 Greg Reinsel		8. CONTRACT OR GRANT NUMBER(s) 15 N00014-76-C-0930
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Carnegie-Mellon University Pittsburgh, Pennsylvania 15213		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 16 NR-042-367
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics and Probability Program Code (436) Arlington, Virginia 22217		12. REPORT DATE 11 June 1976 12 37p.
		13. NUMBER OF PAGES 33
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 14 TR-112, 1-ONR		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Difference equation, autoregressive moving average process, distributed lag models, maximum likelihood estimation, Newton-Raphson method, generalized least squares, asymptotic distribution.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(see reverse side)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

391190

HB

A method is proposed for the estimation of a general class of scalar linear time series models. The model takes the form of a stochastic difference equation for the dependent variable with exogenous variable inputs, and the disturbances are autocorrelated through an autoregressive moving average process. In the present paper an asymptotically efficient yet computationally simple estimation procedure (in the time domain) is derived for this model. The resulting estimator is shown to be asymptotically equivalent to the maximum likelihood estimator and to possess a limiting multivariate normal distribution.